



Psychometric evaluation of the German version of the Brief Reflective Functioning Interview

Sylke Andreas*¹, Paul Plümer¹, Katharina Reichholf¹,
Maria Dehoust², Holger Schulz³, Pia Müllauer¹, Marie G. Rudden⁴,
Birgit Senft⁵, Richard Gaugeler⁶ and Markus Hayden^{1,7}

¹Institut für Psychologie, Universität Klagenfurt, Austria

²Psychosomatische Klinik Ginsterhof, Rosengarten, Germany

³Department of Medical Psychology, University Medical Center Hamburg-Eppendorf, Germany

⁴Berkshire Psychoanalytic Institute, West Stockbridge, Massachusetts, USA

⁵Reha-Klinik für Seelische Gesundheit und Prävention, Klagenfurt, Austria

⁶Öffentliches Krankenhaus Waiern, Feldkirchen i. K., Austria

⁷Klinik Bad Reichenhall, Bad Reichenhall, Germany

Objectives. To measure mentalization in a feasible manner, various instruments have been designed in recent years. The Brief Reflective Functioning Interview (BRFI) is a short interview that is based on the Adult Attachment Interview (AAI). The aim of both studies was to examine the psychometric properties of the German version of the BRFI and to compare them to those of the AAI.

Methods. In Study 1, we examined 60 students using the BRFI and the AAI. In Study 2, the validity of the BRFI was examined using a mixed sample of students and patients ($N = 149$). Trained coders evaluated the Reflective Functioning Scale (RFS) for the BRFI and the AAI.

Results. We found a significant positive correlation between the RFS total scores of the BRFI and those of the AAI. In addition, both interviews showed excellent internal consistency. We could also show that persons with mental disorders exhibit lower levels of RF score than mentally stable individuals. Women had higher RF scores in the BRFI than men in both samples. Persons whose mentalization capacity was rated below average in either the BRFI or the AAI also reported significantly lower mentalization ability in the self-assessment ($p < .01$).

Conclusions. Our results revealed that the RFS scores measured by the BRFI are highly comparable to those measured by the AAI. Our findings support the results of previous studies, suggesting that the BRFI is a reliable, valid and easy-to-administer alternative to the AAI.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

*Correspondence should be addressed to Sylke Andreas, Department of Clinical Psychology, Psychotherapy and Psychoanalysis, Institut für Psychologie, Universität Klagenfurt, Universitätsstr. 65-67, A-9020 Klagenfurt am Wörthersee, Austria (email: sylke.andreas@aau.at).

Practitioner points

- The German version of the Brief Reflective Functioning Interview (BRFI) proved to be a reliable and valid instrument for the assessment of reflective functioning that is shorter in terms of time to complete and the evaluation process than the measurement via the Adult Attachment Interview (AAI).
- Reflective functioning is negatively associated with psychopathology.

Mentalizing is defined as the capacity to understand the intentional or inner mental states of other persons, taking into account one's own intentional states (e.g., beliefs, thoughts, feelings, desires or goals) (Fonagy, 1991; Fonagy & Bateman, 2008). In recent years, a new concept, Mentalization-Based Treatment (MBT), in psychodynamic psychotherapy and research has become increasingly popular (Bateman & Fonagy, 1999, 2009, 2016; 2019; Fonagy, 1998). The objective of MBT is to improve the mentalization capacity of patients with severe personality disorders (Bateman & Fonagy, 2019).

The mentalizing approach was developed by Fonagy, Steele, Morgan, Steele, and Higgitt (1991) within the London-Parent-Child project. The authors observed that a secure attachment of the child is not only dependent on the mother's attachment security (Fonagy, Steele, & Steele, 1991) but also results from the mother's insight to acquire a psychological understanding of her own early childhood relationship with her parents (Fonagy, Steele, Morgan, et al., 1991). With regard to their predominantly positive attachment experiences and the associated secure attachment representations, 79% of these women also had securely attached children compared to 28% of mothers with an insecure attachment style. In particular, it could be confirmed that mothers who reported more negative childhood experiences during interviews were nevertheless able to provide coherent statements and coping strategies (Fonagy, Steele, Morgan, et al., 1991).

To measure mentalization, Fonagy, Target, Steele, and Steele (1998) developed the Reflective Functioning Scale (RFS) based on the Adult Attachment Interview (AAI, George, Kaplan, & Main, 2012) and examined the relationship between attachment and interaction behaviour between mothers and their infants (Bateman & Fonagy, 2019; Fonagy, 1991). The studies demonstrated that women who were more likely to be securely attached achieved higher values on the RFS than mothers with an avoiding attachment behaviour (Fonagy, Steele, & Steele, 1991). The empirical findings of their study made Fonagy, Steele, and Steele (1991); Fonagy, Steele, Morgan, et al. (1991) suggest that there must be a connection between early attachment experiences, self-representation and affect regulation even in patients with severe personality disorders (Bateman & Fonagy, 2019). Empirical studies have demonstrated that patients with severe personality disorders have lower mentalizing capacities than individuals without psychopathology (Fischer-Kern et al., 2010; Fonagy, 2004; Fonagy & Target, 2000; Skarderud, 2007a, 2007b, 2007c). This has also been shown for patients with other mental disorders such as depression (Lemma, Target, & Fonagy, 2011; Markowitz & Meehan, 2009; Staun, Kessler, Buchheim, Kachele, & Taubner, 2010; Taubner, Kessler, Buchheim, Kachele, & Staun, 2011) or eating disorders (Robinson et al., 2016). Furthermore, contemporary research has revealed associations between impairments in mentalization and other psychological phenomena, such as difficulties in social interactions (Fonagy, Luyten, & Bateman, 2015; Hayden, Müllauer, Gaugeler, Senft, & Andreas, 2018), or the use of maladaptive defence patterns (Hayden et al., 2021).

Since the development of the concept of mentalization, various instruments have been developed for the structured assessment (Luyten, Malcorps, Fonagy, & Ensink, 2019). It is possible to measure mentalization through self-report, expert/observer-rated assessment, interviews, questionnaires or experimental tasks. Currently, one of the most

widespread and commonly used instruments for measuring mentalization is the RFS based on the AAI (Fonagy, Steele, & Steele, 1991; Luyten et al., 2019). However, the primary objective of the AAI (George et al., 2012) is to identify the attachment style of a person, which is why not all questions of the AAI are relevant for the assessment of the interviewee's capacity to mentalize (Taubner et al., 2013). Taubner et al. (2013) examined the internal structure and reliability of the RFS in $N = 196$ subjects. They were able to show that those questions that were supposed to provoke a mentalization in the AAI (George et al., 2012) (demand questions) were included in the calculation of the overall score. However, they were able to show that raters also use the information from those questions that do not prompt mentalizing behaviour (permit questions) (George et al., 2012) to form the overall score. In this respect, focussing only on the demand questions may lead to a systematic underestimation of the reflective functioning (RF) score (Taubner et al., 2013).

Since the AAI (George et al., 2012) has an average interview duration of 1–2 hrs plus approximately 8 hrs of transcription time, Rudden, Milrod, Target, Ackerman, and Graf (2006) developed a more specific interview, the Brief Reflective Functioning Interview (BRFI). The questions in the BRFI are intended to provoke mentalization on attachment experiences and were developed on the basis of the AAI. The BRFI consists of ten questions (e.g., 'Tell me something about one of your parents – how is your father or mother?' or 'What do you think, why did your father/mother become what he/she is now?'). Rudden et al. (2006) used seven of the eight demand questions from the AAI and added two more questions on current relationships. The assessment of mentalization is based on the same principle as for the AAI (see RFS, Fonagy et al., 1998), but in contrast to the RFS based on the AAI, each question is rated as a demand question. Rutiman and Meehan (2012) examined the validity of the BRFI in comparison to the AAI in $N = 27$ students. They found an average shorter interview duration of $M = 24$ min for the BRFI in contrast to $M = 44$ min for the AAI. Furthermore, the authors reported a significant positive correlation ($r = .71$) between the RFS scores of the AAI and the BRFI. The raters achieved a good interrater reliability of $ICC = .79$.

Interviews on the assessment of mentalization can be supplemented by self-assessment questionnaires. Hausberg et al. (2012) developed the Mentalization Questionnaire (MZQ) for the economical assessment of mentalization on the basis of people with different mental disorders. The MZQ is primarily used for screening measurements and follow-up assessments of patients undergoing psychotherapeutic treatment. It measures mentalization with 15 items and is presented as a self-evaluation tool. An example item is 'Talking about feelings would mean that they become more and more powerful'. Self-reports and expert-rated instruments such as the AAI (Fonagy et al., 1998) have different advantages and disadvantages (Andreas et al., 2007). While self-reports focus on the perspective of the person, interview procedures focus more on an observer's perspective. There are also differences in terms of feasibility. The MZQ (Hausberg et al., 2012), for example, can be completed in 10 min, whereas the AAI including transcription and rating takes about 10 hrs. Self-report and expert-rated interviews for measuring reflective functioning always take on complementary, mutually reinforcing functions (Andreas et al., 2007). There are several studies that evaluated the psychometric properties of the MZQ (Hausberg et al., 2012; Innamorati et al., 2017; Ponti, Stefanini, Gori, & Smorti, 2019). However, no study has investigated the relationship between expert-rated RFS scores of the AAI or the BRFI and self-reported mentalization.

The aim of the present study was to examine the psychometric properties of the German version of the BRFI. In addition, reliability and validity of the BRFI in a student

sample (Study 1) as well as a mixed student and clinical sample (Study 2) should also be investigated.

Study 1: Comparison Of The Rfs Scores Measured By The Aai And Those Measured By The Brfi – Psychometric Properties Of The Rfs Of The Brfi

Study 1 aimed to compare the psychometric characteristics of the RFS of the BRFI with those of the AAI. We investigated the distribution characteristics of the RFS items of the BRFI in comparison with those of the AAI. In addition, we aimed to determine the feasibility of the administration of the BRFI in comparison with the AAI. In light of the results of Rutimann's and Meehan's study (2012), we assume that the psychometric properties of the BRFI are comparable to the psychometric properties of the AAI.

Method

Translation process of the BRFI

The BRFI was translated into German by our research group (MH, PM and SyAn), as authorized by the test authors (MR) in 2014. The translation and re-translation followed steps 1 to 6 of the translation guidelines of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) (Wild et al., 2005). In detail, the 10 questions of the BRFI were first translated from English into German by the members of our group (MH, PM, and SyAn). The translation was discussed in our group and modifications were made. A consensus version was then translated back into English by an independent native speaker and presented to the test author (MR) for further review. The test author agreed with the English consensus version, which was then translated back into German.

Design

The design of Study 1 is illustrated in Figure 1. To obtain an appropriate dispersion for determining reflective functioning during the use of the interviews, the MZQ (Hausberg et al., 2012) was used as a screening method for the selection of participants from the total of 187 students who had agreed to participate in the study. Based on the distribution of the MZQ scores, 60 of the 187 students were selected for interviews. With $N = 60$ bachelor's students in psychology and economics at the University of Klagenfurt, Austria, one AAI and one BRFI each were conducted at intervals of approximately one week. To avoid interviewer sequence effects, the sequence between the AAI and the BRFI was varied, that is, in half of the sample, the AAI was conducted first and then the BRFI. To recruit the study

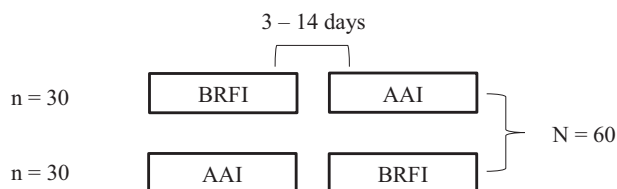


Figure 1. Screening MZQ: low – middle – high scores in self-reported reflective functioning. Note: AAI = Adult Attachment Interview; BRFI = Brief Reflective Functioning Interview; MZQ = Mentalization Questionnaire.

participants, those who were initially considered for the survey were contacted by e-mail either directly in lectures or indirectly via university mailing lists. They were provided with all relevant information on the objectives and the design of the study. Two raters (PP and KR), who had previously been trained by coaches from the Anna Freud Centre in London and achieved good interrater reliabilities ($ICC > .75$) according to Landis and Koch (1977), evaluated the AAI and the BRFIs.

Sample

As mentioned above, to achieve an appropriate distribution of mentalization capacities in the overall sample, the MZQ (Hausberg et al., 2012) was distributed in advance to those students who had agreed to participate in the study. Based on the variance of the total score of the MZQ, the students were assigned to three groups according to the RFS manual by Fonagy et al. (Fonagy et al., 1998): (1) rather low RF ($M \leq 3.46$ – 3.48), (2) average RF and (3) above-average RF ($M \geq 3.9$ – 4.0).

The mean RF of $N = 91$ surveyed psychology students was $M = 3.6$ ($SD = 0.51$). The mean RF in the surveyed sample of economics students ($N = 96$) was comparable to that of psychology students $M = 3.58$ ($SD = 0.5$). To obtain a widely representative sample for the psychometric examination of the BRFI, $N = 60$ students were randomly selected from this population. In each of the three RFS groups, there were $n = 20$ students, of which 25% were economics students and 75% were psychology students. The mean age of the students was 23 years ($SD = 4.6$ years); 48% of the students were female and 52% were male. Nationality was also divided equally between German and Austrian students. Only 2% of the student sample was married, 73% single and 25% with a partner.

Instruments

The Adult Attachment Interview (AAI)

The total of 20 questions of the AAI with their demand format allow the interviewer a semi-structured, clinical process (George et al., 2012). The interview contains questions on the interviewee's life situation in childhood, on the description of significant others, on experiences of loss and how parents reacted when the participant was ill or injured as a child. The aim of the AAI was to provide the attachment style of a person in adulthood and the associated inner working models according to Bowlby (Holmes, 2014). The entire interview is audiotaped for later evaluation purposes to determine the attachment style, the cohesion in the narrative or mentalization.

Based on the questions of the AAI, a scale was developed to measure the RFS (Fonagy et al., 1998). It is the operationalization of the mentalization capacity in the form of a measurement of the subject's capability to reflect about their own mental states and those of others. In particular, the assessment includes the following aspects: The types and frequency of the acknowledgement of the opacity of mental states, empathy and the ability to reflect on certain situations. Furthermore, efforts to combine observable behaviour with inner states of mind are observed and the ability to consider changes in inner states of mind, and the resulting changes in behaviour are assessed. These categories are assessed on an 11-point scale ranging from -1 (anti-reflective) to 9 (exceptionally reflective). Only the odd scores of RFS are provided with anchor examples, where -1 means anti-reflexive, one means missing RF, three means suspicious RF, five means average RF and seven means above-average

RF. However, from a score of four, first reflexive functions are rated. Training is required to administer the RFS.

The Brief Reflective Functioning Interview (BRFI)

The Brief Reflective Functioning Interview (BRFI) (Rudden et al., 2006) is a short, semi-standardized interview. Although it is based on the questions of the AAI, the interviewees are asked only about one parent (Rutimann & Meehan, 2012). Similar to the AAI, the questions of the BRFI are intended to provoke mentalization on attachment experiences. In contrast to the AAI, however, there are fewer questions of episodic memories of attachment experiences. The interviewees are also asked to describe a person of importance in his or her lives to assess the current attachment representation. Rutimann and Meehan (2012) believe that the BRFI captures a general capacity for reflection in relationships rather than a specific attachment experience as in the AAI. As described earlier, the BRFI consists of ten questions. The assessment of mentalization follows the manual for coding mentalization (Fonagy et al., 1998). However, it is not divided into demand and permit questions, but an average score is calculated over all ten questions.

The Mentalization Questionnaire (MZQ)

The MZQ by Hausberg et al. (2012) is a self-assessment instrument for the determination of mentalization in mentally ill persons. The instrument is primarily used for screening examinations, assessments of the course of mental illnesses, and adequate measurements of MBT success. In total, the questionnaire contains 15 different items that can be assigned to the following four subscales and follow a 5-point answer scale (from 1 'do not agree at all' to 5 'fully agree at all'). The reliability of the MZQ has proven to be satisfactory in previous studies, with an internal consistency of $\alpha = .57$ to $.81$. In addition, satisfactory values for test-retest reliability ($r = .54$ to $.76$) were found for both the individual subscales and the total score (Hausberg et al., 2012). In particular, the convergent validity for all external criteria defined in the study (e.g., attachment styles, suicidal tendencies, etc.) showed significant group differences from medium to large effect sizes. Significant correlations between mentalization ability and symptom severity were also determined. Based on the psychometric evaluation and the results presented on the quality criteria, the MZQ can be used as a practical, reliable and valid method for assessments of mentalization from a patient perspective (Hausberg et al., 2012). The instrument has been translated into several languages, including Korean, Indian, Italian, Finnish, and Danish (Juil et al., 2019; Anupama, Bholá, Thirthalli, & Mehta, 2018; Belvederi Murri et al., 2017; Eloranta, Kaltiala, Lindberg, Kaivosoja, & Peltonen, 2020; Ponti et al., 2019; Song & Choi, 2017).

Statistical analyses

Statistical analyses were performed using SPSS version 25. The power calculations were carried out with the programme G*Power 3.1. With a given sample size of $N_{\min} = 50$ persons, an approximate mean effect size according to Cohen (1988) for correlative relationships of $r = .35$ with an alpha error probability of $p < .05$ can be inferentially statistically validated with a sufficiently high power of at least $w = .80$.

The psychometric characteristics for the feasibility and reliability of the RFS of the BRFI were assessed using frequencies of the items of the RF scale and item analyses. We calculated internal consistencies of the RFS of the BRFI in comparison with those of the

AAI. Furthermore, we examined the administration time between both ratings. Moreover, we used a *t* test to investigate the effects of the sequential administration of both interviews. To determine the relationship between the RF scale of the BRFI and the AAI, we performed a Pearson correlation.

Results

Descriptives

The frequencies of the RF BRFI and AAI ratings are shown in Table 1. In both interviews, there were no ratings of -1 (anti-reflective) in the student sample. The students mainly revealed average scores of reflectivity (RF = 3–5). However, we found a score of 9 (extraordinary reflectivity) only for the AAI ($n = 1$).

Distribution characteristics and internal consistency

As presented in Table 2, the means of the RF score of the BRFI were higher for some questions than for others. The highest mean score was revealed for the question ‘current relationship/behaviour’, followed by ‘changes in the relationship’. Other questions such as ‘specific memory’ or ‘reason for chosen the parent’ were more likely to result in lower mean RF scores. The scores of skewness and kurtosis ranged between nearly -1 and 1 .

The internal consistency of the questions of the BRFI was excellent, with a Cronbach’s $\alpha = .92$. The Cronbach’s α for the questions of the AAI was $.91$.

Feasibility of the BRFI

We calculated the administration time to examine the feasibility of the BRFI in comparison with the AAI. On average, the BRFI lasted 26 min ($SD = 9.5$ min), in comparison with the AAI, which lasted 78 min ($SD = 20$ min).

Table 1. Frequencies of the Reflective Functioning (RF) of Brief Reflective Functioning Interview (BRFI) and Adult Attachment Interview (AAI)

Frequencies	BRFI	AAI
RF = -1	0	0
RF = 0	1	0
RF = 1	3	3
RF = 2	4	5
RF = 3	12	9
RF = 4	11	8
RF = 5	13	17
RF = 6	9	7
RF = 7	6	9
RF = 8	1	1
RF = 9	0	1

Note. AAI = Adult Attachment Interview; BRFI = Brief Reflective Functioning Interview; RF = Reflective Functioning.

Table 2. Descriptive statistics of the Reflective Functioning scale (RFS) of the Brief Reflective Functioning Interview (BRFI)

	Mean	SD	Skewness	Kurtosis
Characteristics of a parent	3.70	1.7	-.11	-.63
Reasons behind the characteristics	3.88	1.93	-.03	-1.18
Relationship to parent	3.87	1.82	.01	-.85
Reasons for current relationship	4.18	1.96	-.17	-.68
Specific memory	3.02	1.85	.39	-1.39
Changes over time	4.15	1.84	.04	-.41
Parent's influence/overall experience	3.70	1.88	.63	-.69
Reasons for the choosing parent	3.07	1.50	.30	-.95
Another important person	3.33	1.73	.37	-.91
Relationship to that person	3.63	1.57	.07	-.74

Note. BRFI = Brief Reflective Functioning Interview; RFS = Reflective Functioning scale; SD = standard deviation.

Sequential effects in the interview sequence

We investigated whether there was any difference in the sequence of interviews. In half of the subjects, the AAI was conducted first, and the BRFI was conducted second. In the other half of the subjects, the BRFI was conducted first, and the AAI was conducted second. As depicted in Figure 2, there were no significant differences on the RF scale in terms of the sequential order of the interviews. Furthermore, there was also no significant difference in terms of the length of time between the interviews (3–8 days and 9–14 days, respectively) (see Figure 2).

Comparison between the RF scale of the BRFI and AAI

We found a significant correlation with a large effect size according to Cohen (1988) between the RF scale of the BRFI and that of the AAI ($r = .877, p < .001, n = 60$).

Study 2: Interrater Reliability And Validity Of The Rf Scale Of The Brfi

Study 2 aimed to examine the interrater reliability of the RF scale of the BRFI in the clinical sample and the validity of the RF scale of the BRFI. We investigated the relationship between the expert-rated RF scale and the self-reported scale of the MZQ (Hausberg et al., 2012). To ensure variance in the MZQ responses, we used a mixed sample including the student sample from Study 1 and a clinical sample. Furthermore, we examined age and gender differences in the RF scale of the BRFI between the clinical and the non-clinical sample. We also investigated differences in the RF scores between the clinical and non-clinical sample. We assumed that the reliability should be excellent. We assumed that people who show high reflective functioning in the BRFI will also show high self-reported mentalization in the MZQ. Relating to previous research (Hausberg et al., 2012), we assumed higher scores of mentalization for the non-clinical sample than for the clinical sample. In addition, since persons may differ in age and gender, we explored the differences of the RF scale of the BRFI among the clinical and non-clinical sample.

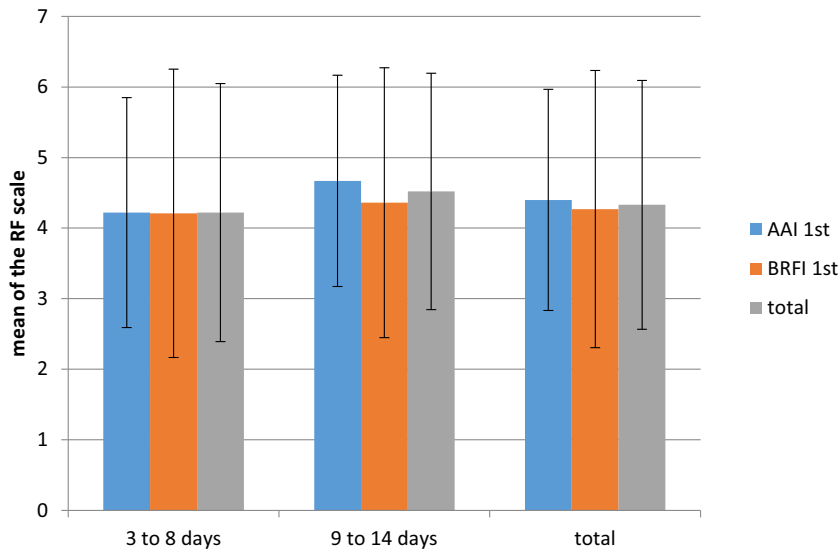


Figure 2. Sequential effect of the interview order. Note: 1st = given first; AAI = Adult Attachment Interview; BRFI = Brief Reflective Functioning Interview; RF scale = Reflective Functioning Scale.

Method

Design

The design of the study is described elsewhere (Hayden, Müllauer, Gaugeler, Senft, & Andreas, 2019). Clinical data were assessed in two different medical centres that use different therapeutic approaches. All patients who met the inclusion criteria were asked to participate at the beginning of treatment. Participation was voluntary, and there were no negative consequences associated with refusal. Regarding inclusion criteria, all participants had to be between 18 and 65 years of age (Hayden et al., 2019). Reasons for exclusion were insufficient ability to understand and/or speak German, early childhood traumatization, acute manic or psychotic episode, dementia, and/or cognitive impairment. All participants provided informed consent that included the purpose and circumstances of the study, information about data privacy, and the right to refuse participation without any consequences, as well as information about the right to withdraw from the study during and after data collection.

Sample

In the clinical sample, $N = 89$ patients were willing to take part in the study (Hayden et al., 2019). The sample description is displayed in Table 3. Most of the participants had diagnoses of the F3 spectrum (mood disorders) and the F4 spectrum (neurotic, stress-related and somatoform disorders) according to ICD-10 (WHO, 1992) as their main diagnosis.

Instruments

The BRFI, AAI and MZQ

The BRFI (Rutimann & Meehan, 2012), the AAI (George et al., 2012) and the MZQ (Hausberg et al., 2012) were already described in Study 1.

Table 3. Sample description of the clinical sample

	Clinic I	Clinic II
n	57	32
Age	M 44.2 (SD 9.92)	M 43.7 (SD 9.70)
Sex		
Male	29	13
Female	28	19
Education		
Elementary	0	1
Main school	12	13
Professional school	8	7
High school	13	4
University	11	3
Other	13	4
Civil status		
Single	11	15
Partnership	11	5
Married	22	6
Divorced	13	5
Widowed	0	1
Children		
Yes	36	20
No	21	12

Note. M = Mean; SD = standard deviation.

Statistics

Statistical analyses were performed using SPSS version 25. To assess the reliability of the RF score of the BRFI in the clinical sample, Intraclass Correlation Coefficients (ICCs) between two trained raters were calculated. Fleiss and Cohen (1973) showed that the ICC is equivalent to weighted kappa for measures of reliability, and Landis and Koch (1977) provided 'rules of thumb' for the interpretation of kappa coefficients. According to these rules, kappa values between 0.21 and 0.40 are 'fair', those between 0.41 and 0.60 are 'moderate', those between 0.61 and 0.80 are 'substantial', and those between 0.81 and 1.00 are 'almost perfect'.

We calculated *t* tests to investigate the relationship between self-reported mentalization capacity and expert-rated mentalization in a mixed sample (student and clinical sample).

Results

Interrater reliability of the BRFI

The raters achieved excellent interrater reliability of $ICC = .85$ for the RFS of the clinical sample.

Validity of the BRFI

There were significant differences in the RFS of the BRFI between the clinical and the non-clinical sample ($p < .001$) (see Table 4). There were also significant gender differences,

since higher RF scores of the BRFI were found for women ($mean\ rank - clinical = 42.22$; $mean\ rank - non-clinical = 72.47$) than for men in both samples ($mean\ rank - clinical = 29.69$; $mean\ rank - non-clinical = 50.44$) ($p < .01$).

There was no significant correlation between the RF score of the BRFI and age in the non-clinical sample ($r = .008$, $p = .95$). However, there was a significant negative correlation between the RF score of the BRFI and age in the clinical sample ($r = -.287$, $p < .05$), indicating that younger patients reveal a higher RF score than older patients.

To further explore the convergent validity, we differentiated the total sample into persons with lower mentalization capacity and persons with average or above-average mentalization capacity ($RF \geq 4$) in the BRFI and AAI in the mixed sample. The t test between the two groups showed a significant difference in self-assessed mentalization ability. Persons who were rated as having below-average mentalization capacity in either the BRFI or the AAI also reported lower mentalization ability in the self-assessment ($M = 3.2$, $SD = .81$) than persons who were rated as having at least an average mentalization capacity in the interviews ($M = 3.49$, $SD = 0.62$) ($p < .01$, $df = 135$).

Discussion

Our studies examined the psychometric properties of the German version of the BRFI in comparison to the AAI and in relation to the self-assessment of mentalization in various samples with a total of $N = 149$ subjects. We showed that the RFS measured by the BRFI is highly comparable to the RFS measured by the AAI. In Study 1, reliability and validity measures were comparable between the BRFI and the AAI, although feasibility was very much in support of using the BRFI. In Study 2, we showed that the RF score of the BRFI was significantly higher in the non-clinical than in the clinical sample. We also found gender differences between the two samples. Regarding the correlation between the RF score in the BRFI and age, we could only demonstrate a negative significant correlation for the clinical sample.

Although there was a high agreement between the two interviews regarding the RFS total scores, the AAI slightly tended to detect persons with exceptional mentalization capacity. The difference was not statistically significant but is certainly of clinical relevance for the use of the instruments. In the questions of the BRFI, as well as in the study by Taubner et al. (2013), there are questions that promote the ability to mentalize, such as the question regarding the change in the relationship to a parent over time. In addition, there are questions where the RF scores are significantly lower on average and do not suggest the ability to mentalize, for example, specific memories of childhood.

Table 4. Descriptive statistics for study 2

	Clinical sample	Non-clinical sample
RFS BRFI		
Mean	2.65	4.33
SD	1.47	1.76
Range	-1 to 5	0 to 8

Note. MZQ = Mentalization Questionnaire; RFS BRFI = Reflective Functioning Scale measured via the Brief Version of the Reflective Functioning Interview.

We also found no indications of sequential effects since it did not matter whether the AAI was given first and then the BRFI or vice versa; there were no significant differences in the RF scores. This is an indication of the independence of the two interviews, which differ slightly in the questions. In addition, both interviews show excellent internal consistencies. It should be noted that both interviews contain questions that are intended to provoke RF, with the BRFI referring less to episodic childhood memories and not containing a question on how to deal with the loss of significant others. Thus, although the BRFI asks for mentalization with different questions, our study shows high similarities, which underlines the validity of the BRFI as detected in the study by Rutimann and Meehan (2012).

Regarding differences in mentalization capability, our results are in line with empirical findings from previous studies (Fischer-Kern et al., 2010, Fonagy, 2004; Fonagy & Target, 2000; Skarderud, 2007a, 2007b, 2007c). These studies could also demonstrate that the mentalizing capacity is higher in mentally stable persons than in patients with mental disorders. This indicates that an impairment in reflective functioning could be an underlying mechanism of psychopathology.

Furthermore, we detected a significant difference between the sexes in the RF score of the BRFI in both the clinical and non-clinical samples. There are previous empirical findings regarding mentalization capacities of men and women, however, these results are heterogeneous. While some studies found no significant differences (Arnott & Meins, 2007; Jessee et al., 2016; Pajulo et al., 2018; Talia et al., 2019; Taubner et al., 2013), other research results confirm our findings (Bouchard et al., 2008; Cologon, Schweitzer, King, & Nolte, 2017). Furthermore, the relationship between mentalization and age has hardly been studied so far. Fonagy et al. (2016) found a small but negative significant correlation between RF scores and age, but only in a non-clinical sample, while other studies could not detect any significant correlation in this regard (Talia et al., 2019; Taubner et al., 2013). In our study, the association was significant in the clinical sample, perhaps due to the variance in age with the same result has been found in the study of Scandurra et al. (2020), who also reported a negative correlation between age and RF scores. The heterogeneity in the findings and the lack of empirical studies indicate a need for further research, especially against the background of gender effects.

To the best of our knowledge, there is currently no other study that has compared self-reported reflective functioning with expert-rated reflective functioning. Although the significant difference between self-rated versus expert-rated reflective functioning is small, it is clinically highly relevant. This demonstrates the concurrent validity of the RF scale of the BRFI.

This, in turn, speaks for the use of the BRFI for research questions in which patients are confronted with many additional questionnaires. The AAI is advantageous if the aim of the study is to capture the attachment style in adulthood and the cohesion of the narrative. It can also be stated critically that slightly higher RF values can be detected in the AAI than in the BRFI. This is probably due to the questions about experiences of loss and to the more comprehensive presentation of the early attachment figures. However, the difference was very small and was not significant.

Our results are comparable with the results of Taubner et al. (2013), reporting that there are no significant differences in the RF score for demand questions coded on the basis of the AAI. The authors conclude that it is, therefore, possible to omit demand questions or exchange them with other questions that provoke RF without affecting the overall score. In contrast to the BRFI, however, Taubner et al. (2013) are in favour of expanding the interview with permit questions because the interviewers probably tend to

include the answers from the permit questions in the overall assessment of the AAI. This probably also explains why in our study, the AAI identified slightly higher values for the total RF score than the BRFI.

Limitations and strengths

In addition to the strengths of this study, which combines an experimental design with a naturalistic design, some limitations should also be mentioned. The first limitation lies in the operationalization or variance of the scale of mentalization. In many studies, it becomes clear that the scale can differentiate well between clinical and non-clinical populations (Taubner et al., 2013), but the variance of the scores is minimal. For example, most values on the RF scale are usually scattered only in the slightly below-average to slightly above-average range with RF scores between 3 and 5 (Hausberg et al., 2012). Considering that the RF scale has a range from –1 to 9, this is problematic with regard to the differentiation of the construct. In this study, too, a sufficient variance in the RF scale could be achieved only by combining the student sample with the clinical sample. There is an urgent need for research and further development with regard to the operationalization of mentalization.

The sample consisted partly of psychology and economics students and one clinical sample of patients from two different clinics. Therefore, our sample cannot be regarded as representative either for a student population or for a clinical population. Apart from these limitations, the results of our studies should be considered satisfactory since the ratings in the AAI and the BRFI followed the RF manual of Fonagy et al. (1998). All the coders were certified by the Anna Freud Centre so that the reliabilities were satisfactory. Furthermore, our study is the first to use a mixed sample to compare the RF rating of a short, economical and valid instrument, the BRFI, with the gold-standard AAI and to arrive at satisfactory results.

Conflict of interest

All authors declare no conflict of interest.

Funding

These studies were funded by the Köhler-Stiftung S112/10102/10.

References

- Andreas, S., Harfst, T., Dirmaier, J., Kawski, S., Koch, U., & Schulz, H. (2007). A psychometric evaluation of the German version of the 'Health of the Nation Outcome Scales, HoNOS-D': On the feasibility and reliability of clinician-performed measurements of severity in patients with mental disorders. *Psychopathology*, *40*, (2), 116–125. <https://doi.org/10.1159/000098492>
- Anupama, V., Bholra, P., Thirthalli, J., & Mehta, U. M. (2018). Pattern of social cognition deficits in individuals with borderline personality disorder. *Asian Journal of Psychiatry*, *33*, 105–112. <https://doi.org/10.1016/j.ajp.2018.03.010>
- Arnott, B., & Meins, E. (2007). Links among antenatal attachment representations, postnatal mind-mindedness, and infant attachment security: A preliminary study of mothers and fathers. *Bulletin of the Menninger Clinic*, *71*, 132–149. <https://doi.org/10.1521/bumc.2007.71.2.132>

- Bateman, A., & Fonagy, P. (1999). Effectiveness of partial hospitalization in the treatment of borderline personality disorder: A randomized controlled trial. *American Journal of Psychiatry*, *156*, 1563–1569. <https://doi.org/10.1176/ajp.156.10.1563>
- Bateman, A., & Fonagy, P. (2009). Randomized controlled trial of outpatient mentalization-based treatment versus structured clinical management for borderline personality disorder. *American Journal of Psychiatry*, *166*(12), 1355–1364. <https://doi.org/10.1176/appi.ajp.2009.09040539>
- Bateman, A., & Fonagy, P. (2016). *Mentalization-based treatment for personality disorders: A practical guide* (1st ed.). Oxford, UK: Oxford University Press.
- Bateman, A., & Fonagy, P. (Eds.) (2019). *Handbook of mentalizing in mental health practice* (2nd ed.). Washington DC: American Psychiatric Association Publishing.
- Belvederi Murri, M., Ferrigno, G., Penati, S., Muzio, C., Piccinini, G., Innamorati, M., . . . Amore, M. (2017). Mentalization and depressive symptoms in a clinical sample of adolescents and young adults. *Child and Adolescent Mental Health*, *22*, 69–76. <https://doi.org/10.1111/camh.12195>
- Bouchard, M.-A., Target, M., Lecours, S., Fonagy, P., Tremblay, L.-M., Schachter, A., & Stein, H. (2008). Mentalization in adult attachment narratives: Reflective functioning, mental states, and affect elaboration compared. *Psychoanalytic Psychology*, *25*(1), 47–66. <https://doi.org/10.1037/0736-9735.25.1.47>
- Cohen, J. (1988). *Statistical power and analysis for the behavioral sciences (second)*. Hillsdale, NJ: Lawrence.
- Cologon, J., Schweitzer, R. D., King, R., & Nolte, T. (2017). Therapist reflective functioning, therapist attachment style and therapist effectiveness. *Administration and Policy in Mental Health*, *44*(5), 614–625. <https://doi.org/10.1007/s10488-017-0790-5>
- Eloranta, S. J., Kaltiala, R., Lindberg, N., Kaivosoja, M., & Peltonen, K. (2020). Validating measurement tools for mentalization, emotion regulation difficulties and identity diffusion among Finnish adolescents. *Nordic Psychology*, *6*(1), 1–23. <https://doi.org/10.1080/19012276.2020.1863852>
- Fischer-Kern, M., Buchheim, A., Hörz, S., Schuster, P., Doering, S., Kapusta, N. D., Taubner, S., Tmej, A., Rentrop, M., Buchheim, P., & Fonagy, P. (2011). "The relationship between personality organization, reflective functioning, and psychiatric classification in borderline personality disorder": Correction to Fischer-Kern et al. (2010). *Psychoanalytic Psychology*, *28*(1), 74–74. <https://doi.org/10.1037/a0022611>
- Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, *33*(3), 613–619. <https://doi.org/10.1177/001316447303300309>
- Fonagy, P. (1991). Thinking about thinking – some clinical and theoretical considerations in the treatment of a borderline patient. *International Journal of Psycho-Analysis*, *72*, 639–656.
- Fonagy, P. (1998). An attachment theory approach to treatment of the difficult patient. *Bulletin of the Menninger Clinic*, *62*, 147–169.
- Fonagy, P. (2004). *Affect regulation, mentalization, and the development of the self* (1. softcover print). New York, NY: Routledge.
- Fonagy, P., & Bateman, A. (2008). The development of borderline personality disorder—a mentalizing model. *Journal of Personality Disorders*, *22*(1), 4–21. <https://doi.org/10.1521/pedi.2008.22.1.4>
- Fonagy, P., Luyten, P., & Bateman, A. (2015). Translation: Mentalizing as treatment target in borderline personality disorder. *Personality Disorders*, *6*(4), 380–392. <https://doi.org/10.1037/per0000113>
- Fonagy, P., Luyten, P., Moulton-Perkins, A., Lee, Y.-W., Warren, F., Howard, S., . . . Lowyck, B. (2016). Development and validation of a self-report measure of mentalizing: The reflective functioning questionnaire. *PLoS One*, *11*(7), e0158678. <https://doi.org/10.1371/journal.pone.0158678>
- Fonagy, P., Steele, M., Steele, H., Moran, G. S., & Higgitt, A. C. (1991). The capacity for understanding mental states: The reflective self in parent and child and its significance for security of attachment. *Infant Mental Health Journal*, *12*(3), 201–218. [https://doi.org/10.1002/1097-0355\(199123\)12:3<201::aid-imhj2280120307>3.0.co;2-7](https://doi.org/10.1002/1097-0355(199123)12:3<201::aid-imhj2280120307>3.0.co;2-7)

- Fonagy, P., Steele, H., & Steele, M. (1991). Maternal representations of attachment during pregnancy predict the organization of infant-mother attachment at one year of age. *Child Development*, *62*, 891–905. <https://doi.org/10.2307/1131141>
- Fonagy, P., & Target, M. (2000). Playing with reality: Iii. The persistence of dual psychic reality in borderline patients. *The International Journal of Psycho-Analysis*, *81*(Pt 5), 853–873.
- Fonagy, P., Target, M., Steele, H., & Steele, M. (1998). *Reflective Functioning Scale manual*. Unpublished manuscript, London, UK.
- George, C., Kaplan, N., & Main, M. (2012). Adult Attachment Interview. In *Bindung im Erwachsenenalter : ein Handbuch für Forschung und Praxis* (pp. 419–439). Bern, Switzerland: Huber.
- Hausberg, M. C., Schulz, H., Piegler, T., Happach, C. G., Klöpffer, M., Brütt, A. L., Sammet, I., & Andreas, S. (2012). Is a self-rated instrument appropriate to assess mentalization in patients with mental disorders? Development and first validation of the Mentalization Questionnaire (MZQ). *Psychotherapy Research*, *22*(6), 699–709. <https://doi.org/10.1080/10503307.2012.709325>
- Hayden, M. C., Müllauer, P. K., Beyer, K. J. P., Gaugeler, R., Senft, B., Dehoust, M. C., & Andreas, S. (2021). Increasing mentalization to reduce maladaptive defense in patients with mental disorders. *Frontiers in Psychiatry*, *12*, 637915. <https://doi.org/10.3389/fpsy.2021.637915>
- Hayden, M. C., Müllauer, P. K., Gaugeler, R., Senft, B., & Andreas, S. (2018). Improvements in mentalization predict improvements in interpersonal distress in patients with mental disorders. *Journal of Clinical Psychology*, *74*, 2276–2286. <https://doi.org/10.1002/jclp.22673>
- Hayden, M. C., Müllauer, P. K., Gaugeler, R., Senft, B., & Andreas, S. (2019). Mentalization as mediator between adult attachment and interpersonal distress. *Psychopathology*, *52*(1), 10–17. <https://doi.org/10.1159/000496499>
- Holmes, J. (2014). *John Bowlby and attachment theory (Second edition). Makers of modern psychotherapy*. London, UK; New York, NY: Routledge, Taylor & Francis Group.
- Innamorati, M., Imperatori, C., Harnic, D., Erbuto, D., Patitucci, E., Janiri, L., . . . Fabbriatore, M. (2017). Emotion regulation and mentalization in people at risk for food addiction. *Behavioral Medicine*, *43*, 21–30. <https://doi.org/10.1080/08964289.2015.1036831>
- Jessee, A., Mangelsdorf, S. C., Wong, M. S., Schoppe-Sullivan, S. J., & Brown, G. L. (2016). Structure of reflective functioning and adult attachment scales: Overlap and distinctions. *Attachment and Human Development*, *18*(2), 176–187. <https://doi.org/10.1080/14616734.2015.1132240>
- Juul, S., Lunn, S., Poulsen, S., Sørensen, P., Salimi, M., Jakobsen, J. C., . . . Simonsen, S. (2019). Short-term versus long-term mentalization-based therapy for outpatients with subthreshold or diagnosed borderline personality disorder: A protocol for a randomized clinical trial. *Trials*, *20*(1), 196. <https://doi.org/10.1186/s13063-019-3306-7>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174. <https://doi.org/10.2307/2529310>
- Lemna, A., Target, M., & Fonagy, P. (2011). The development of a brief psychodynamic intervention (Dynamic Interpersonal Therapy) and its application to depression: A pilot study. *Psychiatry: Journal for the Study of Interpersonal Processes*, *74*, 41–48. <https://doi.org/10.1521/psyc.2011.74.1.41>
- Luyten, P., Malcorps, S., Fonagy, P., & Ensink, K. (2019). Assessment of mentalizing. In A. Bateman, & P. Fonagy (Eds.), *Handbook of mentalizing in mental health practice* (pp. 37–62). Washington DC: American Psychiatric Association Publishing.
- Markowitz, J. C., & Meehan, K. B. (2009). Does depressive state influence reported attachment status? *The Journal of Clinical Psychiatry*, *70*, 1726. <https://doi.org/10.4088/JCP.09lr05363>. author reply 1726–7
- Pajulo, M., Tolvanen, M., Pyykkönen, N., Karlsson, L., Mayes, L., & Karlsson, H. (2018). Exploring parental mentalization in postnatal phase with a self-report questionnaire (PRFQ): Factor structure, gender differences and association with sociodemographic factors. The Finn Brain Birth Cohort Study. *Psychiatry Research*, *262*, 431–439. <https://doi.org/10.1016/j.psychres.2017.09.020>

- Ponti, L., Stefanini, M. C., Gori, S., & Smorti, M. (2019). The assessment of mentalizing ability in adolescents: The Italian adaptation of the Mentalization Questionnaire (MZQ). *TPM-Testing, Psychometrics, Methodology in Applied Psychology*, *26*, 29–38. <https://doi.org/10.4473/TPM26.1.2>
- Robinson, P., Hellier, J., Barrett, B., Barzdaitiene, D., Bateman, A., Bogaardt, A., . . . Fonagy, P. (2016). The NOURISHED randomised controlled trial comparing mentalisation-based treatment for eating disorders (MBT-ED) with specialist supportive clinical management (SSCM-ED) for patients with eating disorders and symptoms of borderline personality disorder. *Trials*, *17*(1), 549. <https://doi.org/10.1186/s13063-016-1606-8>
- Rudden, M., Milrod, B., Target, M., Ackerman, S., & Graf, E. (2006). Reflective functioning in panic disorder patients: A pilot study. *Journal of the American Psychoanalytic Association*, *54*, 1339–1343. <https://doi.org/10.1177/00030651060540040109>
- Rutimann, D. D., & Meehan, K. B. (2012). Validity of a brief interview for assessing reflective function. *Journal of the American Psychoanalytic Association*, *60*(3), 577–589. <https://doi.org/10.1177/0003065112445616>
- Scandurra, C., Dolce, P., Vitelli, R., Esposito, G., Testa, R. J., Balsam, K. F., & Bochicchio, V. (2020). Mentalizing stigma: Reflective functioning as a protective factor against depression and anxiety in transgender and gender-nonconforming people. *Journal of Clinical Psychology*, *76*, 1613–1630. <https://doi.org/10.1002/jclp.22951>
- Skarderud, F. (2007a). Eating one's words, part I: 'concretised metaphors' and reflective function in anorexia nervosa—an interview study. *European Eating Disorders Review*, *15*, 163–174. <https://doi.org/10.1002/erv.777>
- Skarderud, F. (2007b). Eating one's words: Part III. Mentalisation-based psychotherapy for anorexia nervosa—an outline for a treatment and training manual. *European Eating Disorders Review*, *15*, 323–339. <https://doi.org/10.1002/erv.817>
- Skarderud, F. (2007c). Eating one's words, part II: The embodied mind and reflective function in anorexia nervosa—theory. *European Eating Disorders Review*, *15*, 243–252. <https://doi.org/10.1002/erv.778>
- Song, H., & Choi, H.-A. (2017). Exploration of the Factor Structure of the Mentalization Questionnaire (MZQ) in 16–17-year-old Korean Adolescents. *Korean Journal of Clinical Psychology*, *36*(3), 391–401. <https://doi.org/10.15842/kjcp.2017.36.3.009>
- Staun, L., Kessler, H., Buchheim, A., Kachele, H., & Taubner, S. (2010). Mentalization and chronic depression. *Psychotherapeutics*, *55*, 299–305. <https://doi.org/10.1007/s00278-010-0752-9>
- Talia, A., Miller-Bottomo, M., Katznelson, H., Pedersen, S. H., Steele, H., Schröder, P., . . . Taubner, S. (2019). Mentalizing in the presence of another: Measuring reflective functioning and attachment in the therapy process. *Psychotherapy Research*, *29*(5), 652–665. <https://doi.org/10.1080/10503307.2017.1417651>
- Taubner, S., Horz, S., Fischer-Kern, M., Doering, S., Buchheim, A., & Zimmermann, J. (2013). Internal structure of the Reflective Functioning Scale. *Psychological Assessment*, *25*(1), 127–135. <https://doi.org/10.1037/a0029138>
- Taubner, S., Kessler, H., Buchheim, A., Kachele, H., & Staun, L. (2011). The role of mentalization in the psychoanalytic treatment of chronic depression. *Psychiatry*, *74*, 49–57. <https://doi.org/10.1521/psyc.2011.74.1.49>
- WHO. (1992). *The ICD-10 classification of mental and behavioural disorders. Clinical descriptions and diagnostic guidelines*. Geneva, Switzerland: World Health Organisation.
- Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., & Erikson, P. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported Outcomes (PRO) Measures: Report of the ISPOR task force for translation and cultural adaptation. *Value in Health: the Journal of the International Society for Pharmacoeconomics and Outcomes Research*, *8*, 94–104. <https://doi.org/10.1111/j.1524-4733.2005.04054.x>